Rev.R.Acad.Cienc.Exact.Fis.Nat. (Esp)
Vol. 93, N.º 4, pp 451-455, 1999
Monográfico: Problemas complejos de decisión. II

PROCESO DE DECISIÓN BASADO EN LA FUNCIÓN DE ESPARCIMIENTO

(Distribución exponencial/esparcimiento/índice de Gini/*AMS clasificación:* 62N05)

J. M. FERNÁNDEZ PONCE*, M. T. GÓMEZ GÓMEZ, A. SUÁREZ LLORENS

* *Dpto. de Estadística e I.O. Facultad de Matemáticas, Universidad de Sevilla, 41012-Sevilla.*

RESUMEN

En este artículo, se estudia un proceso de decisión de exponencialidad basado en la función de esparcimiento para observaciones no censuradas. Existen muchos tests en la literatura estadística cuando la distribución de la hipótesis alternativa en N.B.U.E. y, usando métodos de simulación , comprobamos que es más potente que los clásicos propuestos.

1. INTRODUCCIÓN

Kochar y Wiens (1987) estudiaron algunos órdenes parciales entre distribuciones de vida con respecto a sus propiedades de envejecimiento. Sea X una variable aleatoria que representa el tiempo de vida de un sistema o componente con función de distribución (f.D) F y soporte en $[0, \infty)$. Se define la función de vida residual media como: $\mu_F(x) = E(X - x|_{X>x})$. Se dice que F es Nueva Mejor (Peor) que la usada en esperanza, es decir, distribución NBUE (NWUE), si $\mu_F(x) \leqslant (\geqslant) \mu_F(0)\ \forall x \geqslant 0$. Se comprueba fácilmente que la distribución exponencial es NBUE y NWUE debido a la propiedad de falta de memoria. Kochar y Wiens (1987) definieron el siguiente orden parcial entre funciones de distribución:

$$F \overset{NBUE}{\leqslant} G \text{ iff } \frac{\mu_F(F^{-1}(u))}{\mu_G(G^{-1}(u))} \leqslant \frac{\mu_F(F^{-1}(0))}{\mu_G(G^{-1}(0))}$$

En la sección 2, definiremos la función de esparcimiento como un L-funcional generalizado, la cual fue utilizado para caracterizar distribuciones de vida en Fernández Ponce et al. (1998). La función de esparcimiento cuantifica la dispersión por la derecha del *u-ésimo* cuantil de la distribución. El estimador natural de esta función es consistente y asintóticamente normal. Si consideramos solamente las distribuciones de vida con soporte en $[0, \infty)$, la función de esparcimiento caracteriza a la distribución exponencial porque es la única que posee una función de esparcimiento lineal.

En los últimos años se han estudiado diferentes tests para contrastar exponencialidad. En concreto se realiza un estudio basado en el Índice de Gini en el trabajo de Gail y Gastwirth (1978, a). Demostramos en este artículo que el estadístico de Gini no es el más apropiado para realizar un contraste de exponencialidad cuando la hipótesis alternativa es cualquier distribución. Estos autores utilizaron el Índice de Gini de una exponencial, que vale 0.5, para realizar contrastes. El razonamiento es que si el Índice de Gini de cualquier distribución alternativa está próximo a cero o a uno entonce la alternativa no es una exponencial. Sin embargo, si la alternativa no es N.B.U.E. ni N.W.U.E. este razonamiento no es v´alido puesto que su función de esparcimiento cruzaría como mínimo una vez a la función de esparcimiento de la distribución exponencial. Este resultado se demostrará en un teorema de la sección 2.

En la sección 3, proponemos un contraste para la exponencial y estudiamos las propiedades de consistencia y normalidad asintótica del correspondiente estadístico. Asimismo, tabulamos la potencia para diferentes alternativas usando el método de Monte Carlo.

2. LA FUNCIÓN DE ESPARCIMIENTO

Sea X una variable aleatoria con función de distribución F. La función de esparcimiento de X se define como la esperanza de la variable aleatoria $\{X - Q_X(u)^+ = \max\{(X - Q_X(u)), 0\}$, donde $Q_X(u)$ es la función cuantil de X. Fernández Ponce et al. (1998) y Shaked y Santhikumar (1998) demostraron que la función de esparcimiento se puede utilizar para caracterizar a las distribuciones de vida mediante sus propiedades de envejecimiento. Si F es una distribución absolutamente continua entonces podemos obtener un estimador natural de la función de esparcimiento. Básicamente, usamos la descomposición de la función de esparcimiento como un L-funcional. Consecuentemente, si $X_{(1)}, ..., X_{(n)}$ son los estadísticos de orden de una muestra de tamaño n de F, se propone el siguiente estimador:

$$S_n^+(u) = \begin{cases} \frac{1}{n}\sum_{j=i+1}^{n} (n - j + 1)(X_{(j)} - X_{(j-1)}) & \text{si } \frac{i-1}{n} < u \leqslant \frac{i}{n} \quad i = 1, ..., n. \\ 0 & \text{si } \frac{n-1}{n} < 1 \leqslant 1. \end{cases} \tag{1}$$

Teorema 1. *Sea X una variable aleatoria absolutamente continua con función de distribución F, y media finita. Entonces* $S_n^+(u)$ *converge casi seguro a* $S_X^+(u)$.

Demostración. Por definición,

$$S_n^+(u) = (1 - u)\, T_u(F_n),$$

donde $T_u(F_n) = \int_0^1 T_{t,u}(F_n)dK_u(t)$, $K_u(t)$ es la distribución vniforme en $[u, 1]$ y $T_{t,u}(F_n) = Q_n(t) - Q_n(u)$. Denotemos $Q_n(.)$ el estimador natural de la función cuantil de F. Usando que F es absolutamente continua, obtenemos que $Q_n(u)$ converge casi seguro a $Q_X(u)$. Por tanto,

$$T_{t,u}(F_n) \xrightarrow{c.s.} T_{t,u}(F).$$

Consecuentemente, $T_u(F_n) \xrightarrow{c.s.} T_u(F)$. Ahora bien, usando el primer teorema de continuidad (ver Borokov, 1984 pág. 38), demostramos el resultado. □

Corolario 1. *Sea X una variable aleatoria no negativa con función de distribución absolutamente continua F y soporte en* $[0, \infty)$. *Entonces, F sigue una distribución exponencial de media* λ *si, y sólo si, la función de esparcimiento de X es:*

$$S_X^+(u) = \lambda(1 - u).$$

Demostración. Sea X una variable aleatoria con función de distribución $F(x;\lambda) = 1 - e^{-x/\lambda}$, para $x \geqslant 0$. Se comprueba fácilmente que $S_X^+(u) = \lambda(1 - u)$.

Recıprocamente si $S_X^+(u) = \lambda(1 - u)$ entonces $\overset{NBUE}{<}$ $Exp(1)$ y $Exp(1)$ $\overset{NBUE}{<}$ F (ver Fernández Ponce et al., 1998). Así que, F difiere a lo sumo en factor escalar positivo de la distribución exponencial de media 1. Consecuentemente, F es una distribución exponencial. □

Corolario 2. *Sea X una variable aleatoria no negativa con función de distribución F, y con media finita. Sea GI el Índice de Gini de X, entonces*

$$GI = \int_0^1 \frac{S_X^+(u)}{S_X^+(0)}\, du$$

Demostración.

Klefsjö (1984) demostró que $GI = \int_0^1 (1 - W(u))du$, donde $W(u)$ es la transformación normalizada del tiempo total de prueba de F. Se demuestra fácilmente que

$$W(u) = 1 - \frac{S_X^+(u)}{S_X^+(0)}, \quad u \in (0, 1)$$

obteniéndose de forma sencilla el resultado. □

3. CONTRASTE NO PARAMÉTRICO DE EXPONENCIALIDAD

Gail y Gastwirth (1978, a) usaron el Índice de Gini muestral para contrastar exponencialidad. La idea consistía en comparar el área bajo la función de esparcimiento normalizada, es decir $S_X^+(u)/S_X^+(0)$, debido a que el área correspondiente a la exponencial es 0.5. Pero esta comparación es sólo válida para distribuciones NBUE (NWUE) ya que en este caso la función de esparcimiento es:

$$S_X^+(u) \leqslant (\geqslant) \lambda(1 - u) \quad \forall u \in [0, 1],$$

(ver Fernández Ponce et al., 1998). Más aún, existen muchas distribuciones que no cumplen ninguna de las dos propiedades anteriores, y en estos casos, la función de esparcimiento cruza al menos una vez la recta $\lambda(1 - u)$, donde λ es la media de la distribución. Ahora, se propone un contraste alternativo para contrastar exponencialidad basado en la distancia L_1 entre la recta $\lambda(1 - u)$ y la función de esparcimiento normalizada correspondiente a la distribución alternativa.

Supongamos que deseamos realiza el contraste:

$$\begin{aligned} &H_0 : F(x; \lambda) = 1 - e^{-x/\lambda} \quad x > 0,\ \lambda \text{ desconocida} \\ &H_1 : F(.; \lambda) \text{ no es exponencial y con} \\ &\qquad \text{soporte en } [0, \infty). \end{aligned}$$

Usando el Corolario 2, proponemos la siguiente medida de desviación de H_0 respecto a H_1:

$$\Delta(F) = \int_0^1 \left| 1 - u - \frac{S_X^+(u)}{S_X^+(0)} \right| du$$

Fijémonos que $0 \leqslant \Delta(F) \leqslant 1/2$ y $\Delta(F) = 0$ si $F \in H_0$. Por tanto, estimaremos $\Delta(F)$ usando una versión suavizada de (1) basado en la interpolación lineal de una función. Consecuentemente proponemos el siguiente estadístico:

$$\Delta_n(F) = \int_0^1 \left| 1 - u - \frac{\hat{S}_X^+(u)}{S_X^+(0)} \right| du$$

donde:

$$\hat{S}_X^+(u) = \begin{cases} (i - nu)S_X^+(\frac{i-1}{n}) + (nu - i + 1)S_X^+(\frac{i}{n}), & \text{si } \frac{i-1}{n} < u \leqslant \frac{i}{n},\ i = 1, ..., n-1 \\ 0 & c.c. \end{cases}$$

Particularmente, si $X_{(1)}, ..., X_{(n)}$ son los estadístico de orden de una muestra aleatoria de tamaño n procedente de F (suponiendo que $X_{(0)} = 0$), entonces definimos las siguientes variables aleatorias para $i = 1, ..., n$.

$$A_i = \frac{-n(n - i + 1)(X_{(i)} - X_{(i-1)})}{\Sigma_{i=1}^n X_i}$$

$$B_i = \frac{i(n-i+1)(X_{(i)}-X_{(i-1)}) + \Sigma_{j=i+1}^n (n-j+1)(X_{(j)}-X_{(j-1)})}{\Sigma_{i=1}^n}$$

$$M_{i,n} = \begin{bmatrix} 1 - B_i & 1 + A_i \\ i - 0{,}5 & n \end{bmatrix} \quad N_{i,n} = \begin{bmatrix} 1 - B_i & 1 + A_i \\ n^2C_i^2 - i^2 & 2n^2C_i - 2in \end{bmatrix}$$

Por tanto, es fácil demostrar que

$$\Delta_n(F) = \frac{1}{n^2} \sum_{i=1}^n |det(M_{i,n}) + I_{(\infty,\,0)}(D_i)\, det(N_{i,n})|$$

Estudiemos ahora las propiedades asintóticas de $\Delta_n(F)$.

Teorema 2. *Sea X una variable aleatoria no negativa con función de distribución absolutamente continua F, con media finita y soporte en* $[0, \infty)$. *Entonces* $\Delta_n(F)$ *converge casi seguro a* $\Delta(F)$.

Demostración.

En primer lugar, demostremos que $\hat{S}_X^+(u)$ converge casi seguro a $S_X^+(u)$. Por definición, $S_n^+(u - \frac{1}{n}) < \hat{S}_X^+(u) \leqslant S_X^+(u)\ \forall u \in (0, 1)$. Ahora bien, usando el teorema A.1, obtenemos que

$$S_X^+\left(u - \frac{12}{n}\right) \xrightarrow{c.s.} S_X^+(u)$$

Consecuentemente, $\hat{S}_X^+(u) \xrightarrow{c.s.} S_X^+(u)$. Usando el primer teorema de continuidad (ver Borokov, 1984, pág. 38) y el teorema de convergencia dominada, obtenemos el resultado fácilmente. □

Teorema 3. *Si F es una distribución NBUE o NWUE con* $EX^2 < \infty$ *pero no es una distribución exponencial, entonces*

$$\frac{\Delta_n(F) - E\Delta_n(F)}{\sqrt{var(\Delta_n(F))}} \xrightarrow{P} N(0;\ 1).$$

Demostración.

Tenemos que

$$\Delta_n(F) = \int_0^1 \left[1 - u - \frac{\hat{S}_X^+(u)}{S_X^+(0)} \right]^+ du + \int_0^1 \left[1 - u - \frac{\hat{S}_X^+(u)}{S_X^+(0)} \right]^- du$$

Si F es una distribución NBUE (NWUE), entonces

$$\int_0^1 \left[1 - u - \frac{\hat{S}_X^+(u)}{S_X^+(0)} \right]^+ du \quad \text{y} \quad \int_0^1 \left[1 - u - \frac{\hat{S}_X^+(u)}{S_X^+(0)} \right]^- du$$

convergen casi seguro a cero. Consecuentemente, converge en probabilidad a cero. Usando el teorema de Slutzky (ver Chernoff et al., 1967), obtenemos que

$$[\Delta_n(F)] \overset{\mathcal{L}}{\equiv} \int_0^1 \left[1 - u - \frac{\hat{S}_X^+(u)}{S_X^+(0)} \right] du$$

donde, denotamos la misma distribución asintótica por $\overset{\mathcal{L}}{\equiv}$.

Pero se demuestra fácilmente que

$$\int_0^1 \left[1 - u - \frac{\hat{S}^+(u)}{S^+(0)} \right] du = \frac{n-1}{2n} - \frac{n-1}{n} G_n$$

donde G_n es el Índice de Gini. Ahora usando la normalidad asintótica de G_n, demostramos el resultado. □

Fijémonos que si F es una distribución exponencial, tenemos que

$$n\sqrt{12}\left[\Delta_n(F) - \frac{1}{2n} \right] \xrightarrow{\mathcal{L}} N(0, 1).$$

4. MÉTODOS COMPUTACIONALES

En esta sección, tabulamos algunos percentiles de $\Delta_n(F)$ bajo H_0 (ver Tabla 1). Es difícil conseguir la distribución exacta del estadístico bajo la hipótesis nula, sin embargo $\Delta_n(F)$ es un estadístico libre de escala y depende de los espacios normalizados de una distribución exponencial, $D_i = (n - i + 1)(X_{(i)} - X_{(i-1)})$. Hemos generado 500 muestras de tamaño n de una distribución exponencial de media 1. También, hemos comparado la potencia para diferentes alternativas del estadístico $\Delta_n(F)$ frente a

Tabla 1. Valores críticos para $\Delta_n \Lambda(F)$

n	$\Delta_n(F)$ 0,05	0,1	0,9	0,95	0,99
3	0,0311	0,0438	0,2152	0,2409	0,2955
4	0,0400	0,0504	0,2074	0,2569	0,2916
5	0,0412	0,0530	0,1887	0,2258	0,2835
10	0,0332	0,0394	0,1537	0,1782	0,2268
15	0,0324	0,0380	0,1215	0,1460	0,1926
20	0,0278	0,0315	0,1112	0,1308	0,1701
25	0,0245	0,0291	0,0953	0,1143	0,1322
30	0,0255	0,0289	0,0899	0,1040	0,1273
35	0,0237	0,0279	0,0802	0,0899	0,1154
40	0,0224	0,0261	0,0744	0,0902	0,1133

G_n y $L_n(0,5)$ para $n = 20$, donde G_n es el estadístico de Gini (ver Gail y Gastwirth 1978 a) y $L_n(0,5)$ es el estadístico muestral de Lorenz (ver Gail y Gastwirth 1978 b). La potencia ha sido tabulada para 1000 muestras de tamaño $n = 20$ (ver Tabla 2). Las muestras aleatorias para cada alternativa se han obtenido usando el lenguaje de simulación Simscript en ordenadores VAX. La distribución de Pareto fue generada con la transformación inversa de la probabilidad. En la Tabla 2, podemos observar que $\Delta_{20}(F)$ es mejor que G_{20} y $L_{20}(0,5)$ para alternativas NWUE (la distribución Weibull con parámetro de forma 0,8)), la distribución Gamma con parámetro de forma 0,5 y la distribución Pareto para $k = 3$). También se ha comparado G_n con $\Delta_n(F)$ para mixturas en la distribución alternativa (ver Tabla 3). En esta tabla, usamos un test bilateral para G_n y $L_n(0,5)$, hemos generado 1000 muestras de tamaño 20 para la mixtura entre la bistribución Weibull con parámetro de forma 0,8 y una distribución Weibull con parámetro de forma 1,5.

APÉNDICE

Teorema A.1. *Sea $Q_X(u)$ la única solución a $F_X(x^-) \leqslant u \leqslant F_X(x)$, entonces*

$$Q_n(u - a_n) \xrightarrow{c.s.} Q_X(u)$$

Tabla 3. Estimación de la potencia para alternativas $\varepsilon W(0,8) + (1 - \varepsilon)W(1,5)$

ε	$\Delta_{20}(F)$	G_{20}	$L_{20}(0,5)$
0,1	0,325	0,466	0,330
0,25	0,328	0,215	0,201
0,5	0,489	0,109	0,091
0,75	0,572	0,126	0,101
0,9	0,662	0,185	0,156

para toda sucesión decreciente no negativa $\{a_n : n \in \mathbb{N}\}$ con límite cero y donde $Q_n(.)$ es el estimador natural de la función cuantil.

Demostración.

Sea $\varepsilon > 0$. Por la condición de unicidad y la definición de $Q_X(u - a_n)$, tenemos que

$$F_X[Q_X(u - a_n) - \varepsilon] < u - a_n < F_X[Q_X(u - a_n) + \varepsilon] \ \forall n \in N$$

También sabemos que

$$F_n(Q_X(u - a_n) - \varepsilon) \xrightarrow{c.s.} F_X(Q_X(u) - \varepsilon)$$
$$F_n(Q_X(u - a_n) + \varepsilon) \xrightarrow{c.s.} F_X(Q_X(u) + \varepsilon)$$

Consecuentemente,

$$P\{F_m(Q_X(u - a_m) - \varepsilon) < u - a_m < F_m(Q_X(u - a_m) + \varepsilon)\} \to 1$$

Es decir,

$$P\{Q_X((u - a_m) - \varepsilon < Q_m(u - a_m) < Q_X(u - a_m) + \varepsilon)\} \to 1$$

Entonces,

$$P\{\sup_{m \geqslant n} |Q_m(u - a_m) - Q_X(u - a_m)| > \varepsilon\} \to 0, \quad n \to \infty$$

A partir de aquí es fácil conseguir el resultado deseado. □

Tabla 2. Estimación de la potencia para alternativas $\alpha = 0,05$

Alternativas	$\Delta_2^0(F)$	G_{20} unilateral	G_{20} bilateral	$L_{20}(0,5)$ unilateral	$L_{20}(0,5)$ bilateral
Weibull, forma = 0,8	0,723	0,336	0,239	0,320	0,216
Pareto, $k = 3$	0,822	0,553	0,470	0,468	0,337
Gamma, forma = 0,55	0,897	0,565	0,566	0,695	0,603

REFERENCIAS

1. Borokov, A. A. (1984). *Estadística Matemática.* Editorial Mir Moscú.
2. Chernoff, H., Gastwirth, J. L. and Johns, M. V. (1967). Asymptotic distribution of linear combination of functions of order statistics with applications to estimation. *Annals of Statistics* **38,** 52-72.
3. Fernández-Ponce, J. M., Kochar, S. C. and Muñoz-Pérez, J. (1998). Partial orderings of distributions based on right-spread functions. *Journal of Applied Probability* **35,** 221-228.
4. Gail, M. H. and Gastwirth, J. L. (1978, a). A scale-free goodness of fit test for the exponential distribution based on the Gini statistic. *J.R. Stat Soc Ser B* **40,** 3: 350-357.
5. Gail, M. H. and Gastwirth, J. A. (1978, b). A scale-free goodness of fit test for the exponential distribution based on the Lorenz curve. *J.A.S.A.* **37,** 364:787-793.
6. Klesfjö, B. (1984). Reliability interpretations of some concepts from economics. *Naval Research Logistic Quarterly* **31,** 301-308.
7. Kochar, S. C. and W*iens, D. P. (1987). Partial orderings of life distributions with respect to their aging properties. *Naval Research Logistic Quarterly* **34,** 823-829.
8. Shaked, M. and Shanthikumar, J. G. Two variability orders. *Prob Eng Infor Sci* **12,** 1-23.